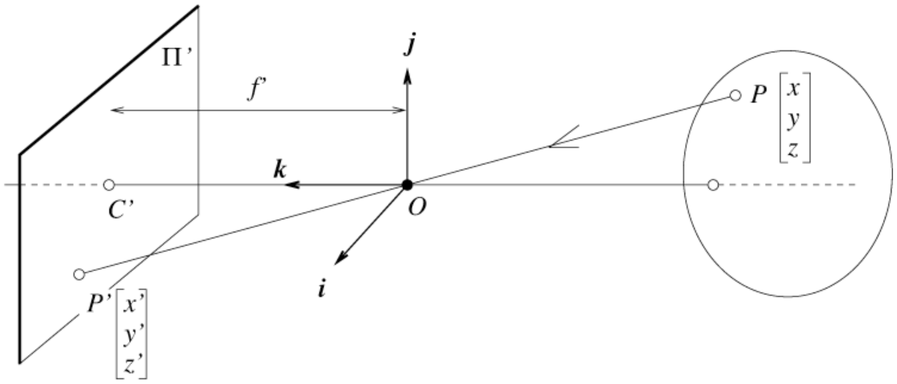
3D Geometry

Our Plan

- Projection from 2D to 3D
- Representing 3D pose
- Projective Invariants.
  - Affine invariants (scaled orthographic projection of planar objects).
  - Projective invariants (planar, perspective).
  - Lack of invariants for 3D objects.

The equation of projection



(Forsyth & Ponce)

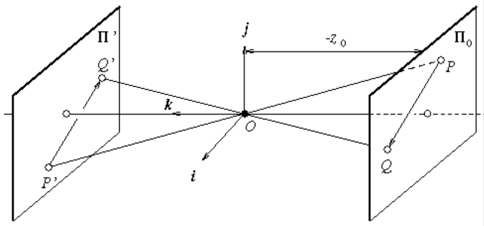
The equation of projection

- Cartesian coordinates:
  - We have, by similar triangles, that (x, y, z) -> (f x/z, f y/z, -f)
  - Ignore the third coordinate, and get

$$(x, y, z) \rightarrow (f\frac{x}{z}, f\frac{y}{z})$$

Weak perspective (scaled orthographic projection)

- Issue
  - perspective effects, but not over the scale of individual objects
  - collect points into a group at about the same depth, then divide each point by the depth of its group



(Forsyth & Ponce)

The Equation of Weak Perspective

$$(x, y, z) \to s(x, y)$$

• *s* is constant for all points.

Projection

- We'll talk about a fixed camera, and moving object.
- Key point:

Points

$$P = \begin{pmatrix} x_1 & x_2 & & & & x_n \\ y_1 & y_2 & . & . & . & y_n \\ z_1 & z_2 & & & & z_n \\ 1 & 1 & & & & 1 \end{pmatrix}$$

Some matrix

$$S = \begin{pmatrix} s_{1,1} & s_{1,2} & s_{1,3} & t_x \\ s_{2,1} & s_{2,2} & s_{2,3} & t_y \end{pmatrix}$$

The image

$$I = \begin{pmatrix} u_1 & u_2 & . & . & . & u_n \\ v_1 & v_2 & & & & v_n \end{pmatrix}$$

Then: $I = SP$

Rotation

$$\begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{pmatrix} P$$

Represents a 3D rotation of the points in P.

First, look at 2D rotation (easier)

Matrix R acts on points by rotating them.

$$R = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 & x_2 & . & . & . & x_n \\ y_1 & y_2 & & & & y_n \end{pmatrix}$$

• Also, RR^T = Identity. R^T is also a rotation matrix, in the opposite direction to R.

Why does multiplying points by R rotate them?

• Think of the rows of R as a new coordinate system. Taking inner products of each points with these expresses that point in that coordinate system.

- • This means rows of R must be orthonormal vectors (orthogonal unit vectors).

• Think of what happens to the points (1,0) and (0,1). They go to (cos theta, -sin theta), and (sin theta, cos theta). They remain orthonormal, and rotate clockwise by theta.

- • Any other point, (a,b) can be thought of as a(1,0) + b(0,1). R(a(1,0)+b(0,1) = Ra(1,0) + Ra(0,1) = aR(1,0) + bR(0,1). So it's in the same position relative to the rotated coordinates that it was in before rotation relative to the x, y coordinates. That is, it's rotated.

Simple 3D Rotation

$$\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 & x_2 & . & . & . & x_n \\ y_1 & y_2 & & & & y_n \\ z_1 & z_2 & & & & z_n \end{pmatrix}$$

Rotation about z axis.

Rotates x,y coordinates. Leaves z coordinates fixed.

Full 3D Rotation

$$R = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix}$$

• Any rotation can be expressed as combination of three rotations about three axes.

$$RR^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

• Rows (and columns) of *R* are orthonormal vectors.

• R has determinant 1 (not -1).

• Intuitively, it makes sense that 3D rotations can be expressed as 3 separate rotations about fixed axes. Rotations have 3 degrees of freedom; two describe an axis of rotation, and one the amount.

• Rotations preserve the length of a vector, and the angle between two vectors. Therefore, (1,0,0), (0,1,0), (0,0,1) must be orthonormal after rotation. After rotation, they are the three columns of R. So these columns must be orthonormal vectors for R to be a rotation. Similarly, if they are orthonormal vectors (with determinant 1) R will have the effect of rotating (1,0,0), (0,1,0), (0,0,1). Same reasoning as 2D tells us all other points rotate too.

- Note if R has determinant -1, then R is a rotation plus a reflection.

Full 3D Motion/Projection

Scale, Projection, 3D Translation, 3D Rotation:

$$s\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}\begin{pmatrix}1 & 0 & 0 & t_x\\ 0 & 1 & 0 & t_y\\ 0 & 0 & 1 & t_z\end{pmatrix}\begin{pmatrix}r_{1,1} & r_{1,2} & r_{1,3} & 0\\ r_{2,1} & r_{2,2} & r_{2,3} & 0\\ r_{3,1} & r_{3,2} & r_{3,3} & 0\\ 0 & 0 & 0 & 1\end{pmatrix}P$$

$$\equiv\begin{pmatrix}s_{1,1} & s_{1,2} & s_{1,3} & st_x\\ s_{2,1} & s_{2,2} & s_{2,3} & st_y\end{pmatrix}P$$

where

$$(s_{1,1}, s_{1,2}, s_{1,3}) \bullet (s_{2,1}, s_{2,2}, s_{2,3}) = 0$$

$$\|(s_{1,1}, s_{1,2}, s_{1,3})\| = \|(s_{2,1}, s_{2,2}, s_{2,3})\|$$

We can just write st_x as t_x and st_y as t_y.

Definitions and Invariants

- A definition of a class means that given a list of properties:
  - For all props, all objects have that prop.
  - No other objects have all properties
- Invariant is an image property that:
  - For some objects, property is true for all images.
  - For all other objects, property is false for all images.

Whiteboard

- Definitions are composed of invariant properties.

Invariants, a brief history

- Invariance has long history in perception.

  Each movement we make by which we alter the appearance of objects should be thought of as an experiment desgned to test whether we have understood correctly the invariant relations of the phenomena before us, that is, their existence in definite spatial relations.

  – Helmholtz, 1878

  If invariants of the energy flux at the receptors of an organism exist, and if these invariants correspond to the permanent properties of the environment … then I think thee is new support for … a new theory of perception in psychology.

  – Gibson, 1967

- In math, Erlanger program conceives geometry as study of invariant properties under a group of transformations.

Invariants on a line

- WLOG, line is *y=0,z=0*.

$$\begin{pmatrix} u_1 & u_2 & . & . & . & u_n \\ v_1 & v_2 & & & & v_n \end{pmatrix} = \begin{pmatrix} s_{1,1} & s_{1,2} & s_{1,3} & t_x \\ s_{2,1} & s_{2,2} & s_{2,3} & t_y \end{pmatrix} \begin{pmatrix} x_1 & x_2 & . & . & . & x_n \\ 0 & 0 & & & & 0 \\ 0 & 0 & & & & 0 \\ 1 & 1 & & & & 1 \end{pmatrix}$$

$$= \begin{pmatrix} s_{1,1} & t_x \\ s_{2,1} & t_y \end{pmatrix} \begin{pmatrix} x_1 & x_2 & . & . & . & x_n \\ 1 & 1 & & & & 1 \end{pmatrix}$$

Then, we can show that ||p3-p2||/||p2-p1|| is invariant to this transformation.

whiteboard

||(s1x3+tx,s2x3+ty)-(s1x2+tx,s2x2+ty)||

/||(s1x2+tx,s2x2+ty)-(s1x1+tx,s2x1+ty)||

= ||(s1(x3-x2),s2(x3-x2))||/||(s1(x2-x1),s2(x2-x1))||

=sqrt(s1^2+s2^2)(x3-x2)/sqrt(s1^2+s2^2)(x2-x1)

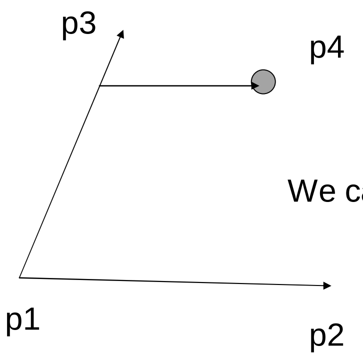
=(x3-x2)/(x2-x1)

Planar Invariants

$$\begin{pmatrix} u_1 & u_2 & . & . & . & u_n \\ v_1 & v_2 & & & & v_n \end{pmatrix} = \begin{pmatrix} s_{1,1} & s_{1,2} & s_{1,3} & t_x \\ s_{2,1} & s_{2,2} & s_{2,3} & t_y \end{pmatrix} \begin{pmatrix} x_1 & x_2 & . & . & . & x_n \\ y_1 & y_2 & & & & y_n \\ 0 & 0 & & & & 0 \\ 1 & 1 & & & & 1 \end{pmatrix}$$

$$= \begin{pmatrix} s_{1,1} & s_{1,2} & t_x \\ s_{2,1} & s_{2,2} & t_y \end{pmatrix} \begin{pmatrix} x_1 & x_2 & . & . & . & x_n \\ y_1 & y_2 & & & & y_n \\ 1 & 1 & & & & 1 \end{pmatrix}$$

A t



We call (a,b) *affine coordinates* of p4.

p4 = p1 + a(p2-p1) + b(p3-p1)

A(p4)+ t = A(p1+a(p2-p1) + b(p3-p1)) + t

= A(p1)+t + a(A(p2)+t – A(p1)-t) + b(A(p3)+t – A(p1)-t)

p4 is linear combination of p1,p2,p3. Transformed p4 is same linear combination of transformed p1, p2, p3.

Perspective Projection

- Problem: perspective is non-linear.
- Solution: Homogenous coordinates.
  - Represent points in plane as *(x,y,w)*
  - *(x,y,w), (kx, ky, kw), (x/w, y/w, 1)* represent same point.
  - If we think of these as points in 3D, they lie on a line through origin. Set of 3D points that project to same 2D point.

Perspective motion and projection

$$\begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \end{pmatrix} \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & 0 \\ r_{2,1} & r_{2,2} & r_{2,3} & 0 \\ r_{3,1} & r_{3,2} & r_{3,3} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \equiv \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_x \\ r_{2,1} & r_{2,2} & r_{2,3} & t_y \\ r_{3,1} & r_{3,2} & r_{3,3} & t_z \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

For Planar Objects

$$\begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_x \\ r_{2,1} & r_{2,2} & r_{2,3} & t_y \\ r_{3,1} & r_{3,2} & r_{3,3} & t_z \end{pmatrix} \begin{pmatrix} x \\ y \\ 0 \\ 1 \end{pmatrix} \equiv \begin{pmatrix} r_{1,1} & r_{1,2} & t_x \\ r_{2,1} & r_{2,2} & t_y \\ r_{3,1} & r_{3,2} & t_z \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

The first two columns on right are orthonormal. Scale is irrelevant. So there are 6 degrees of freedom.

We ignore constraints to get 8. This is called a **projective transformation.**

Projective Transformations

- Mapping from plane to plane.
- Form a group.
  - They can be composed
  - They have inverses.
  - Projective transformations equivalent to set of images of images.

Planar Projective Invariants

- Strategy.
  - Suppose P represents five points. V1 transforms P so that first 4 to canonical position, and fifth to (a,b,c).
  - Next, suppose we are given TP, with T unknown. Find V2 to transform first 4 points of TP to canonical position.
  - V2 = V1*T^{-1}. V2P has fifth point = (a,b,c).
  - For this to work, V1, V2 must be uniquely determined.

Transform to Canonical Position

Example: transform point 1 to (0,0,1). Three linear equations with 8+1 unknowns.

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = k_1 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Similarly, transform other points to: (1,0,1), (0,1,1), (1,1,1). Get 12 equations, 4 unknowns.

- Unique solution.
- Must be non-degenerate. This will be true if no three points collinear.

Affine

- Affine transformations are a subgroup of projective, with last row = (0,0,1).
- Note that this is equivalent to what we did in the affine case. Affine coordinates are coordinates of 4^{th} point after first three are transformed to (0,0), (1,0), (0,1).

Cross Ratio

- Let p1, p2, p3, p4 be collinear points.
- Let (xj,yj) denote the coordinates of pj.
- Let |xj xk| denote the determinate of a matrix whose first column is xj1, xj2, and whose second column is xk1, xk2.
- Cross(p1,p2,p3,p4) =

  (|x1 x2| |x3 x4|)/(|x1 x3| |x2 x4|)
- This cross-ratio is invariant to projective transformations.

Lines: Parameterization

- Equation for line: $ax+by+c=0$.
- Parameterize line as $l = (a,b,c)^{T}$.
- $p=(x,y,1)^T$ is on line if <p,l>=0.

Line Intersection

- The intersection of l and l’ is l x l’ (where x denotes the cross product).
- This follows from the fact that the cross product is orthogonal to both lines.

Intersection of Parallel Lines

- Suppose l and l' are parallel. We can write l=(a,b,c), l' = (a,b,c'). l x l' = (c'-c)(b,-a,0). This equivalent to (b,-a,0).
- This point corresponds to a line through the focal point that doesn't intersect the image plane.
- We can think of the real plane as points (a,b,c) where c isn't equal to 0. When c = 0, we say these points lie on the ideal line at infinity.
- Note that a projective transformation can map this to another line, the horizon, which we see.

Invariants of Lines

- Notice that affine transformations are the subgroup of projective transformations in which the last row is (0, 0, 1).
- These map the line at infinity to itself.
- So parallel lines are affine invariants, since they continue to intersect at infinity.

Invariance in 3D to 2D

- Invariance isn't captured by mathematical definition of invariance because 3D to 2D transformations don't form a group.
  - You can't compose or invert them.
- Let f be a function on images. f is an invariant iff for every Object O, if I1 and I2 are images of O, f(I1)=f(I2).
- f is a non-trivial invariant if there exist two images I1 and I2 such that f(I1)~=f(I2).

Non-Invariance in 3D to 2D

- Theorem: Valid objects are any 3D point sets of size k, for some k. There are no non-trivial invariants of the images of these objects under perspective projection.

Proof Strategy

- Let f be an invariant.
- Suppose two objects, A and B have a common image. Then f(I)=f(J) if I and J are images of either A or B.
- Given any O0, Ok, we construct a series of objects, O1, …, Ok, so that Oi and O(i+1) have a common image for all i, and Ok and j have a common image.
- So for any pair of images, I, J, from any two objects, f(I) = f(J).

Constructing O1 … Ok-1

- Oi has its first i points identical to the first i points of Ok, and the remaining points identical to the remaining points of O0.
- If two objects are identical except for one point, they produce the same image when viewed along a line joining those two points.
  - Along that line, those two points look the same.
  - The remaining points always look the same.

Visibility

- Three geometric steps of determining appearance are projection, pose, and visibility.
- Two issues
  - Is surface point facing the camera
    - Inner product between outward normal and vector to camera center must be >0.
  - Is anything blocking it from view.

Recognition by synthesis

- Blanz and Vetter give example, optimize pose to fit rendered model and image.
- Z-buffer algorithm for visibility.
  - Buffer is image size, set all z values to infinity.
  - Project each triangle into image.
  - For each pixel center it contains, update the z value if this triangle is closer, and render it into image
    - If this triangle has larger z value, don't render it.
- Often special hardware for this.

Preprocess model to determine consistent feature sets.

- Example: if an object has 20 features, might recognize it by looking for all 20 in image.
- Suppose "object" is a sheet of paper with ten features on each side. We should represent this as two sets of ten features, not one set of 20.

Aspect Graphs

- Assume an image of an object is described by a set of features.
- An aspect is a (maximal) set of views of the object that all contain the same features.
- In feature matching, should only match features that share an aspect.
- For complex objects, # of aspects becomes very large.

Aspect Graphs

- Simple example, convex polyhedra.
  - If a side is facing away from the camera (for orthographic projection, the z component of the surface normal is +) it is not visible.
  - Otherwise, the side is completely visible. (only true due to convexity).
  - So if a side has surface normal, n, and we express the viewing direction with a vector v, the side is visible iff <n,v> > 0. (Lhs means inner product).
  - Thus, the boundaries between qualitatively different views are planes in the 3D space of viewing directions.
    - Each plane goes through the origin.
    - Intersect a unit sphere in great circles.
    - Create 2(n choose 2) vertices and O(n^2) cells on the unit sphere. That is, O(n^2) aspects.

Why don't we use Aspect Graphs

- Real (non-convex, non-polyhedral) objects have large aspect graphs.
  - Hard to represent and to compute.
- Many aspects may be tiny, not worth representing. Coarse aspect graph might be worthwhile, but the idea is not well developed.